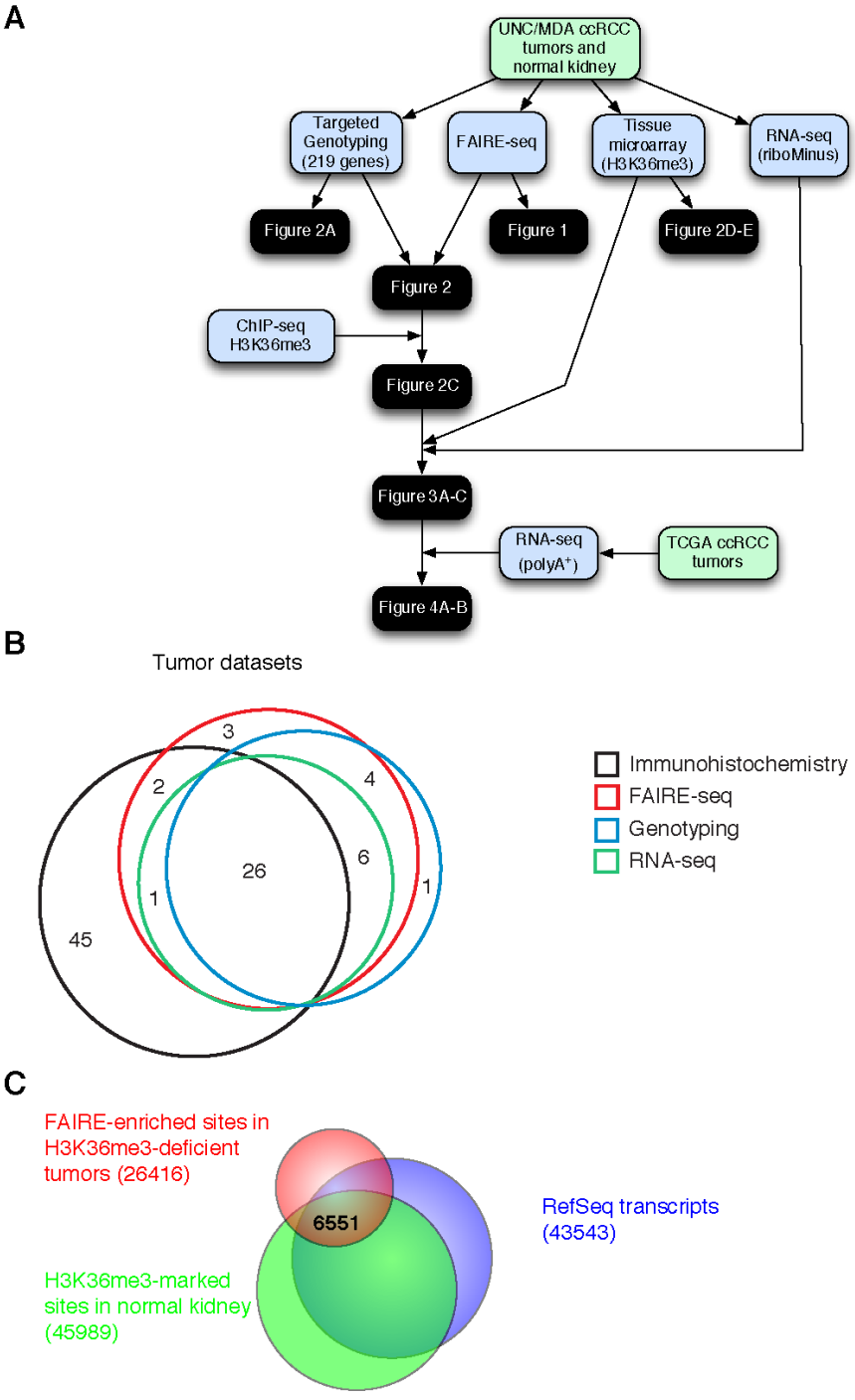
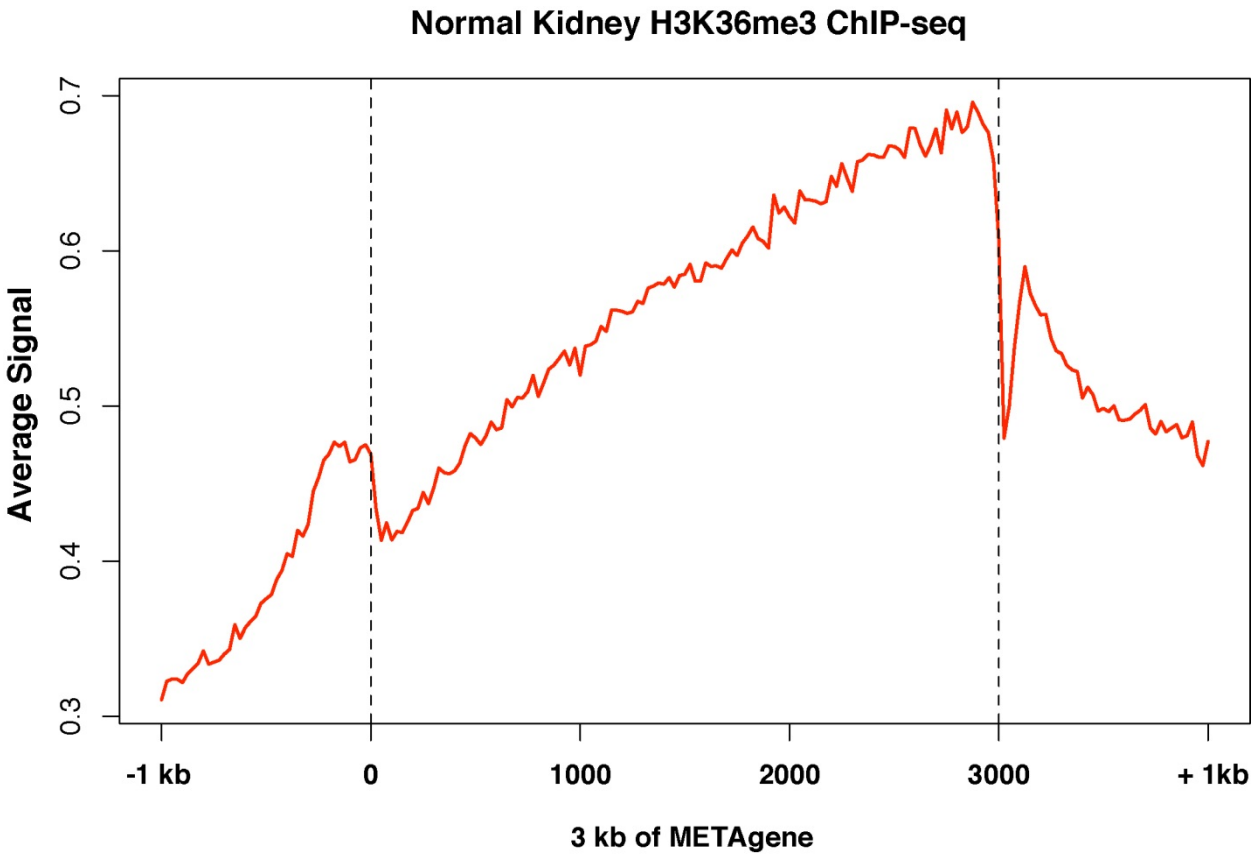
Supplementary Figures and Legends



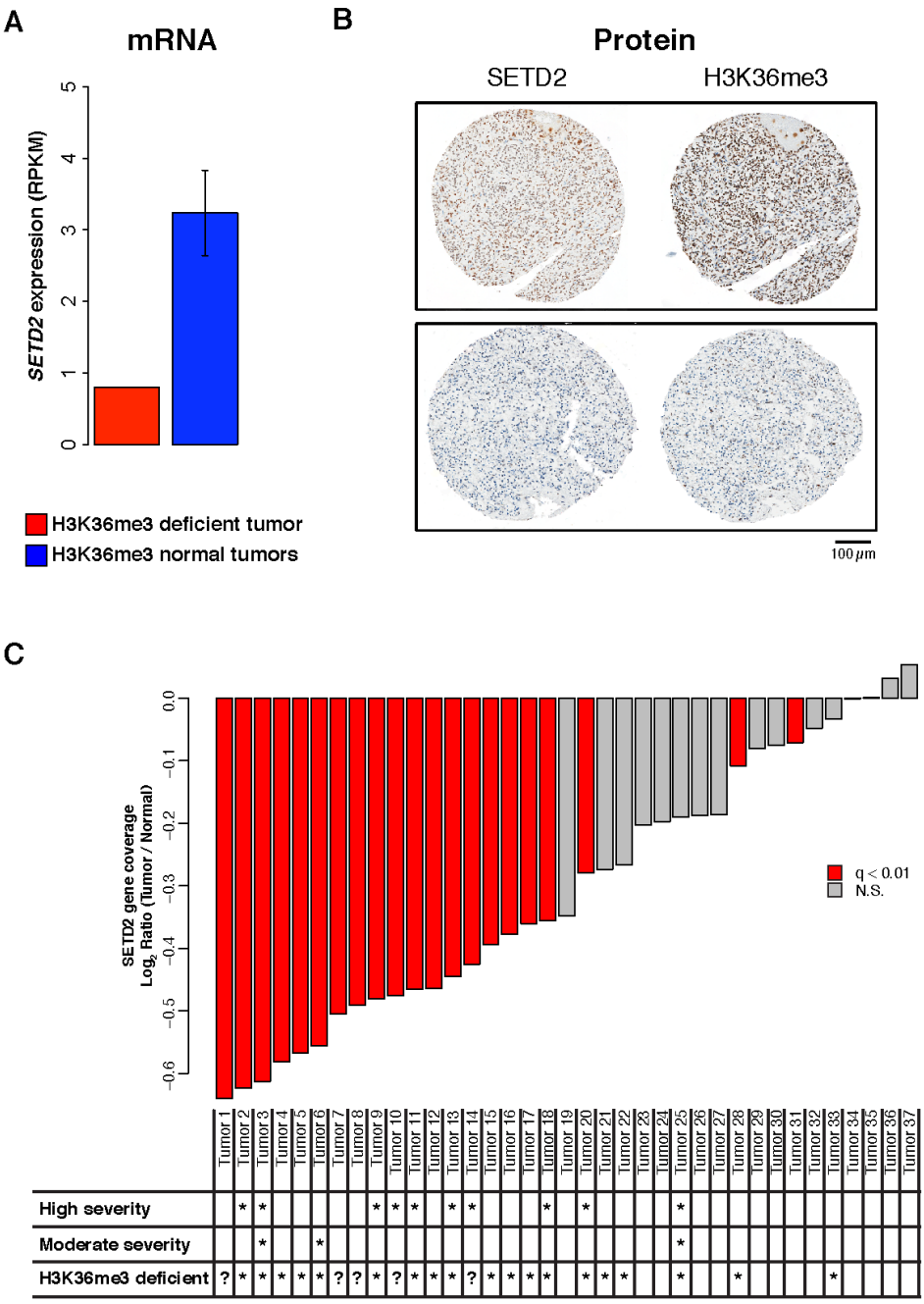
Supplementary Figure 1. Schematic representation of dataset integration and genomic site identification. **A.** Flowchart depicting dataset integration utilized for each figure. Datasets are colored in green, data types in blue, and resulting figures in black. **B.** Venn diagram depicting how tumors were utilized for various experimental approaches. **C**. Venn diagram depicting the intersection of the RefSeq transcripts, H3K36me3-marked regions and genes with FAIRE enrichment in H3K36me3-deficient tumors relative to H3K36me3-normal tumors to yield the 6551 genomic sites used for determination of intron retention.

	Ontology	Term Name	Hyper FDR Q-Val
Cluster 1	MSigDB Perturbation	Genes up-regulated in MCF7 cells (breast cancer) under hypoxia conditions.	1.72E-11
	MSigDB Perturbation	Genes up-regulated in response to both hypoxia and overexpression of an active form of HIF1A [Gene ID=3091].	5.07E-11
	MSigDB Perturbation	Genes up-regulated in MCF7 cells (breast cancer) treated with hypoxia mimetic DMOG [PubChem=3080614].	5.93E-10
	Pathway Commons	HIF-1-alpha transcription factor network	1.51E-08
	MSigDB Perturbation	Genes down-regulated in MCF7 cells (breast cancer) after knockdown of both HIF1A and HIF2A [Gene ID=3091, 2034] by RNAi.	2.28E-07
	MSigDB Perturbation	Genes up-regulated in MCF7 cells (breast cancer) after stimulation with NRG1 [Gene ID=3084].	3.48E-07
	Mouse Phenotype	respiratory system inflammation	1.36931E-06
	Disease Ontology	neck neoplasm	1.52881E-06
	Disease Ontology	neck cancer	1.79918E-06
	Pathway Commons	Hypoxic and oxygen homeostasis regulation of HIF-1-alpha	2.61673E-06
	PANTHER Pathway	PDGF signaling pathway	2.76948E-06
	MSigDB Perturbation	Genes down-regulated in MCF7 cells (breast cancer) after knockdown of HIF1A [Gene ID=3091] by RNAi.	1.262E-05
	Mouse Phenotype	lung inflammation	2.15588E-05
	Mouse Phenotype	abnormal kidney excretion	2.26746E-05
	MSigDB Perturbation	Genes down-regulated by MYC [Gene ID=4609], according to the MYC Target Gene Database.	3.20939E-05
	Disease Ontology	neoplasm of body of uterus	3.62718E-05
Cluster 2	MSigDB Perturbation	Genes within amplicon 17q21-q25 identified in a copy number alterations study of 191 breast tumor samples.	6.85E-13
Cluster 3	None significant to $q < 1 \times 10^{-5}$		

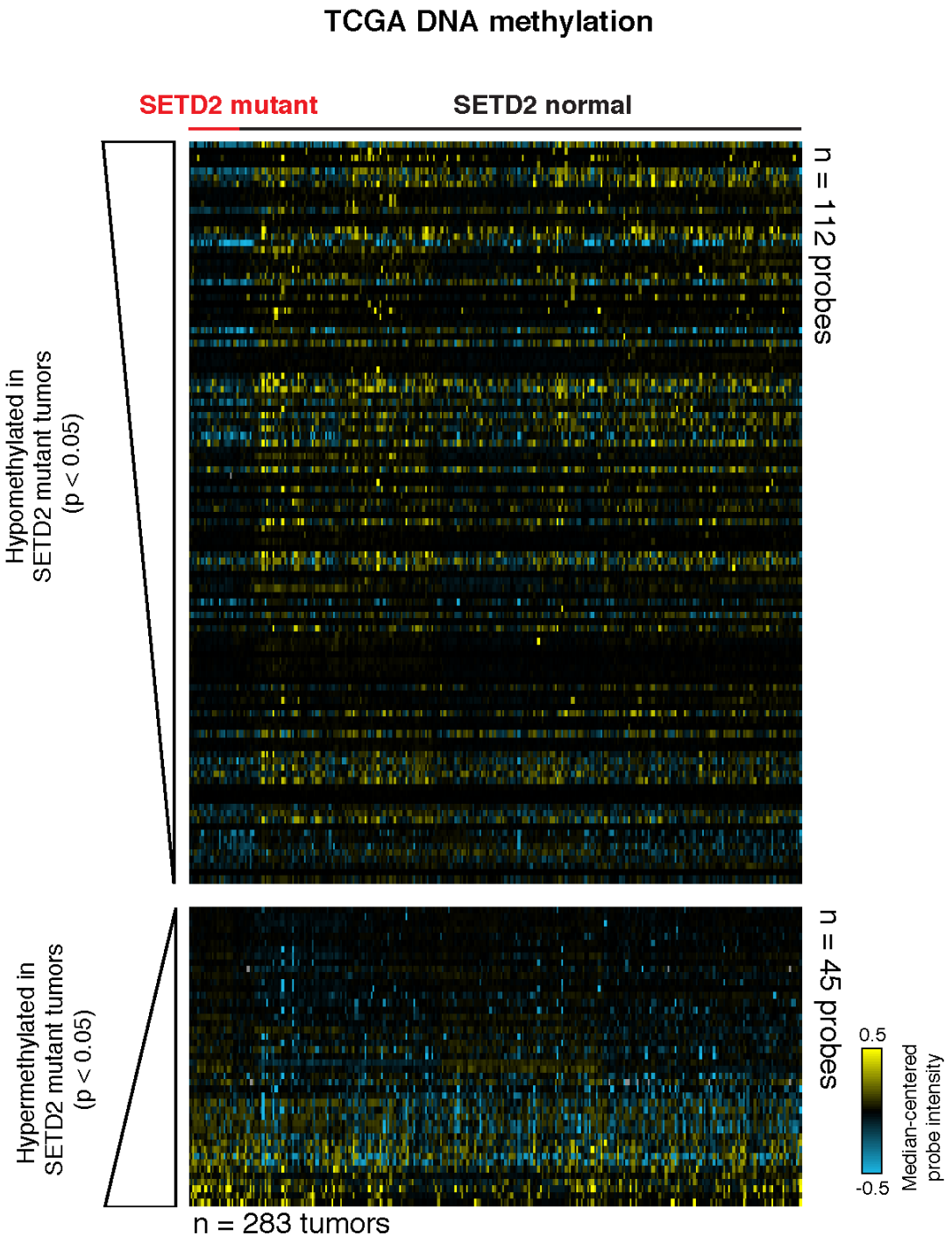
Supplementary Figure 2. Gene ontology associations with sites in Clusters 1-3. The Genomic Regions Enrichment of Annotations Tool (GREAT[6]) was used to analyze the functional significance of regions identified by FAIRE. Associations with hypergeometric FDR-adjusted q-values less than 1×10^{-5} are shown. Cluster 3 analysis yielded no ontologies that met this threshold.



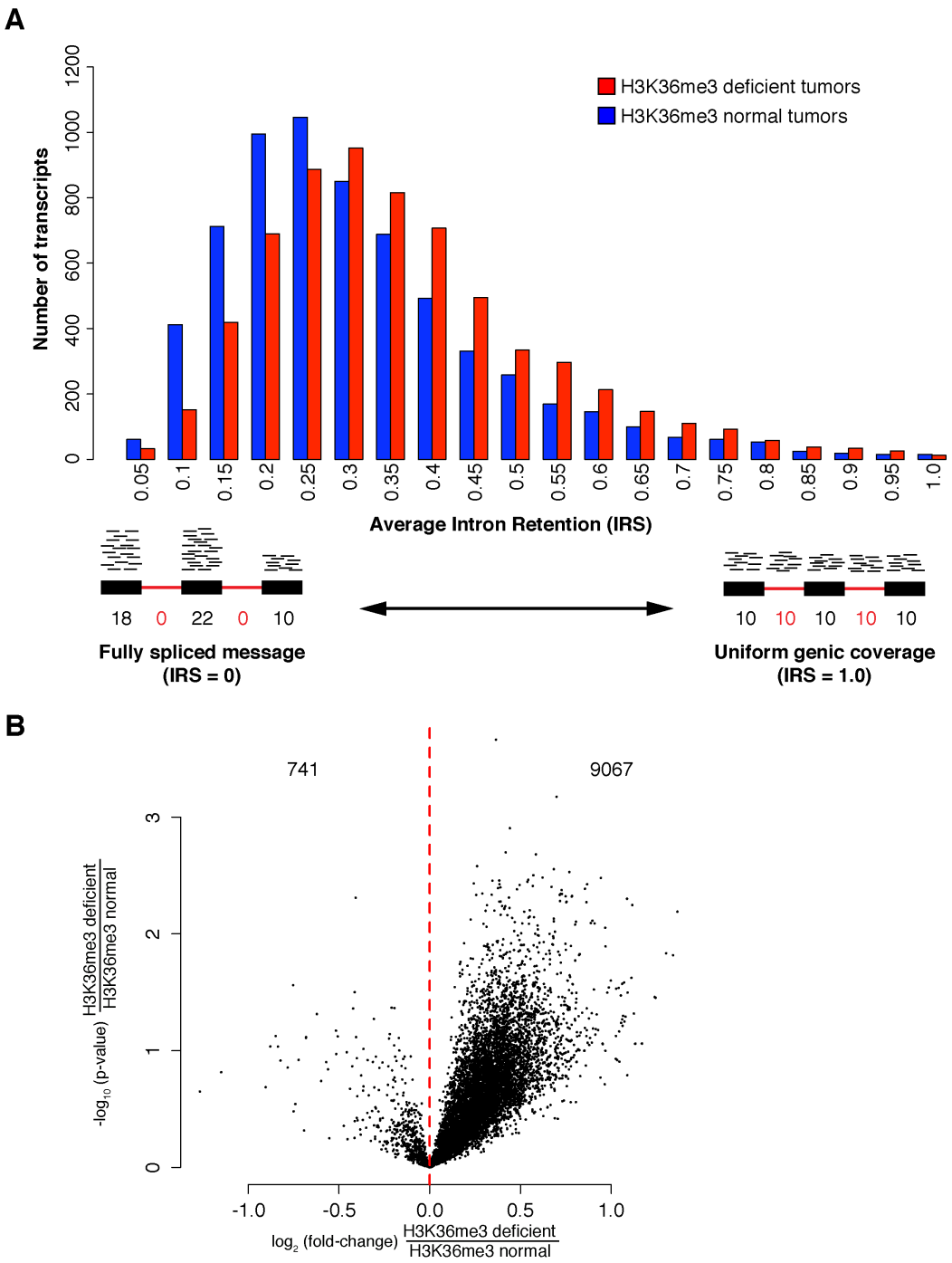
Supplementary Figure 3. **META-gene plot of H3K36me3 ChIP-seq signal from normal kidney.** The average number of reads is plotted across the 3 kb of average gene length, plus 1 kb upstream and downstream, demonstrating a 3' bias for accumulation of the H3K36me3 mark.



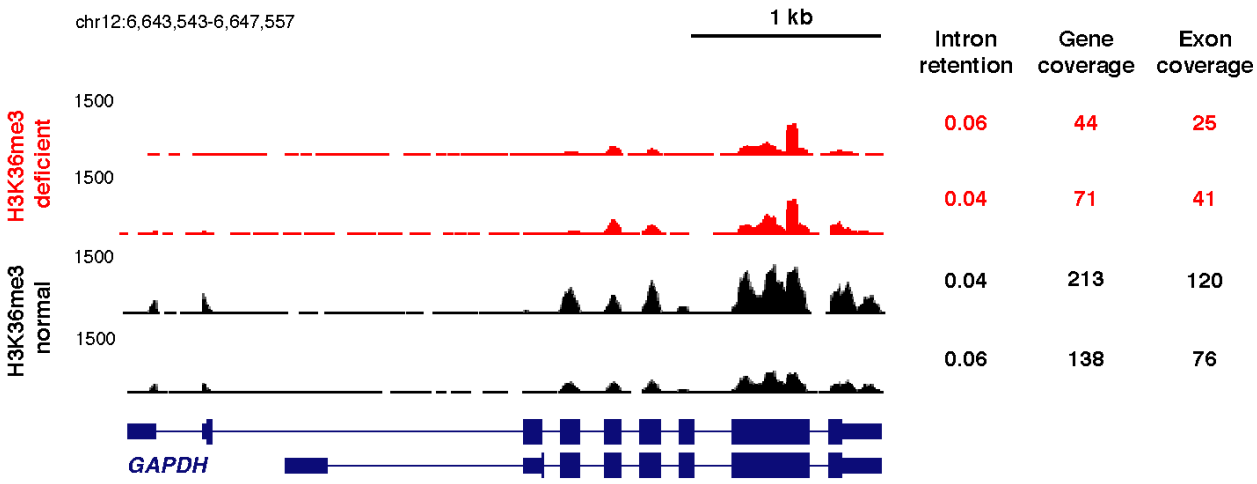
Supplementary Figure 4. Decreased SETD2 expression in *SETD2*-normal tumors results in H3K36me3 deficiency. **A.** SETD2 RNA expression (RPKM) for an H3K36me3-deficient tumor without *SETD2* mutation compared to the average RPKM for H3K36me3-normal tumors. Error bars represent standard error. **B.** Representative immunohistochemical staining of SETD2 protein and H3K36me3 in a genotypically *SETD2*-normal tumor with H3K36me3 deficiency (bottom panel) compared to a *SETD2*- normal tumor with normal SETD2 protein and H3K36me3 levels (top panel). **C.** Log-ratio of gene coverage of each tumor over the average of two normal kidney samples following a log-transformation and mean-centering of the number of reads mapping to each gene. Significance of the tumor-normal difference was determined using a negative binomial test. H3K36me3 status was unknown for 5 tumors (denoted by "?").



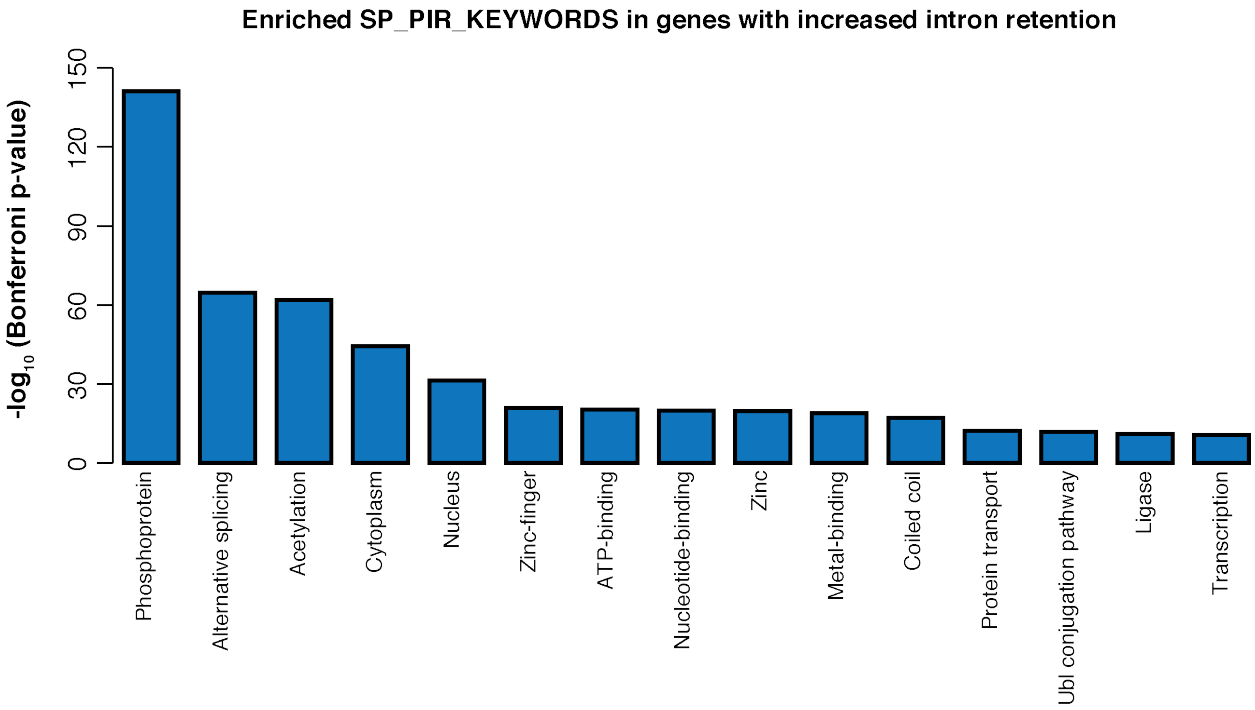
Supplementary Figure 5. Nucleosome-depleted regions in *SETD2*-mutant tumors display localized DNA hypomethylation. Median-centered DNA methylation intensity for probes in genes both displaying FAIRE enrichment associated with *SETD2* mutation and marked by H3K36me3 in normal kidney. Data from 283 TCGA ccRCC tumors at 157 probes are presented. Specific hypomethylation of a cohort of these regions was selectively associated with SETD2 mutation. Color was assigned on a scale of -0.5 to 0.5.



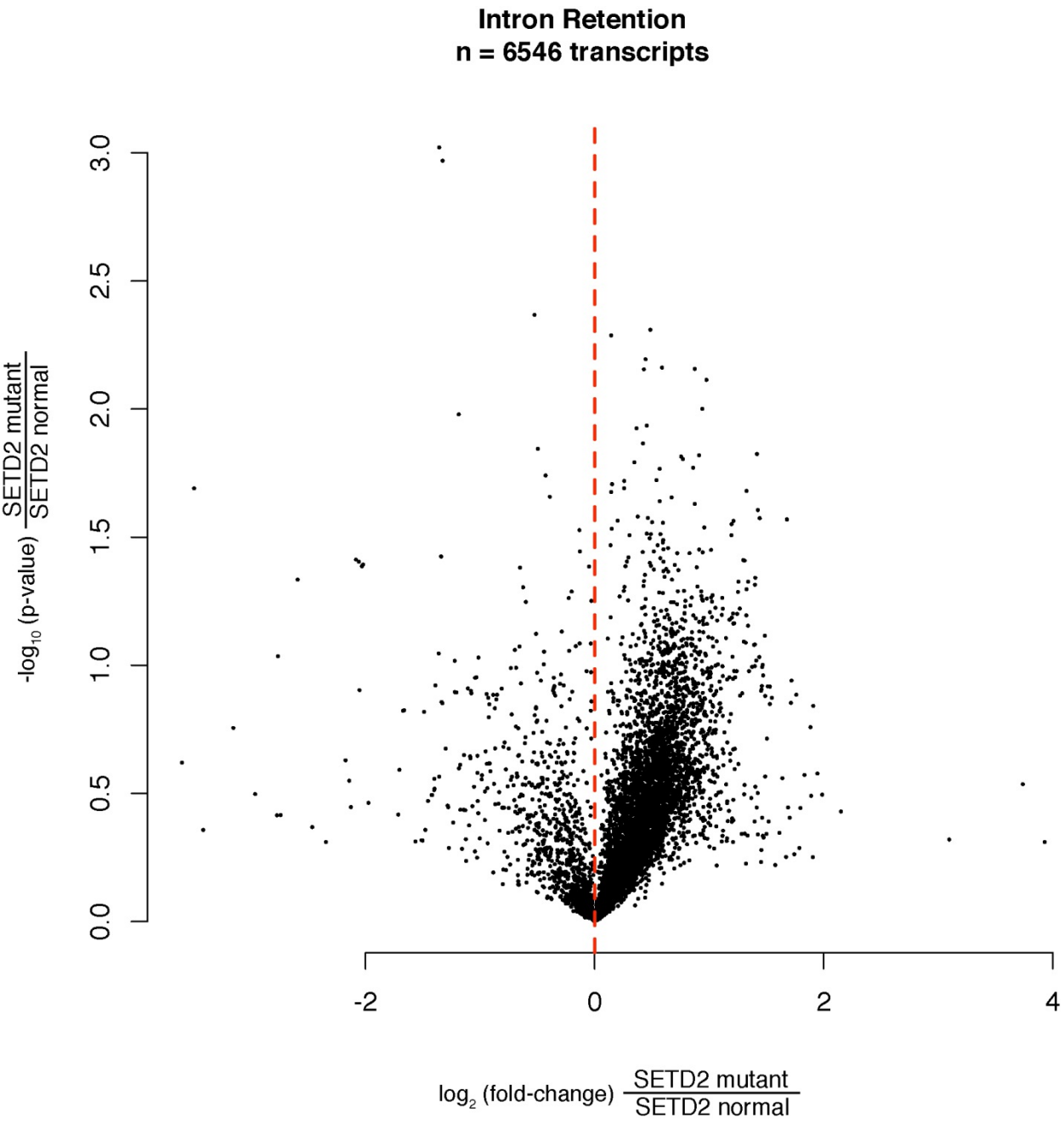
Supplementary Figure 6. H3K36me3-deficient tumors display increased intron retention compared to H3K36me3-normal tumors. A. The number of RefSeq transcripts is plotted for each gradation of Intron Retention Score (IRS) in rRNA-depleted RNA. An average intronic retention score of 0 indicates exclusively exonic coverage (fully spliced) whereas an intronic retention score of 1.0 indicates uniform genic coverage (the absence of exonic enrichment). **B.** Intron Retention Scores for selected genes (**Fig. S1C**, however H3K36me3 ChIP-seq data was obtained from normal kidney of a second individual) were compared between H3K36me3-deficient tumors and H3K36me3-normal tumors.



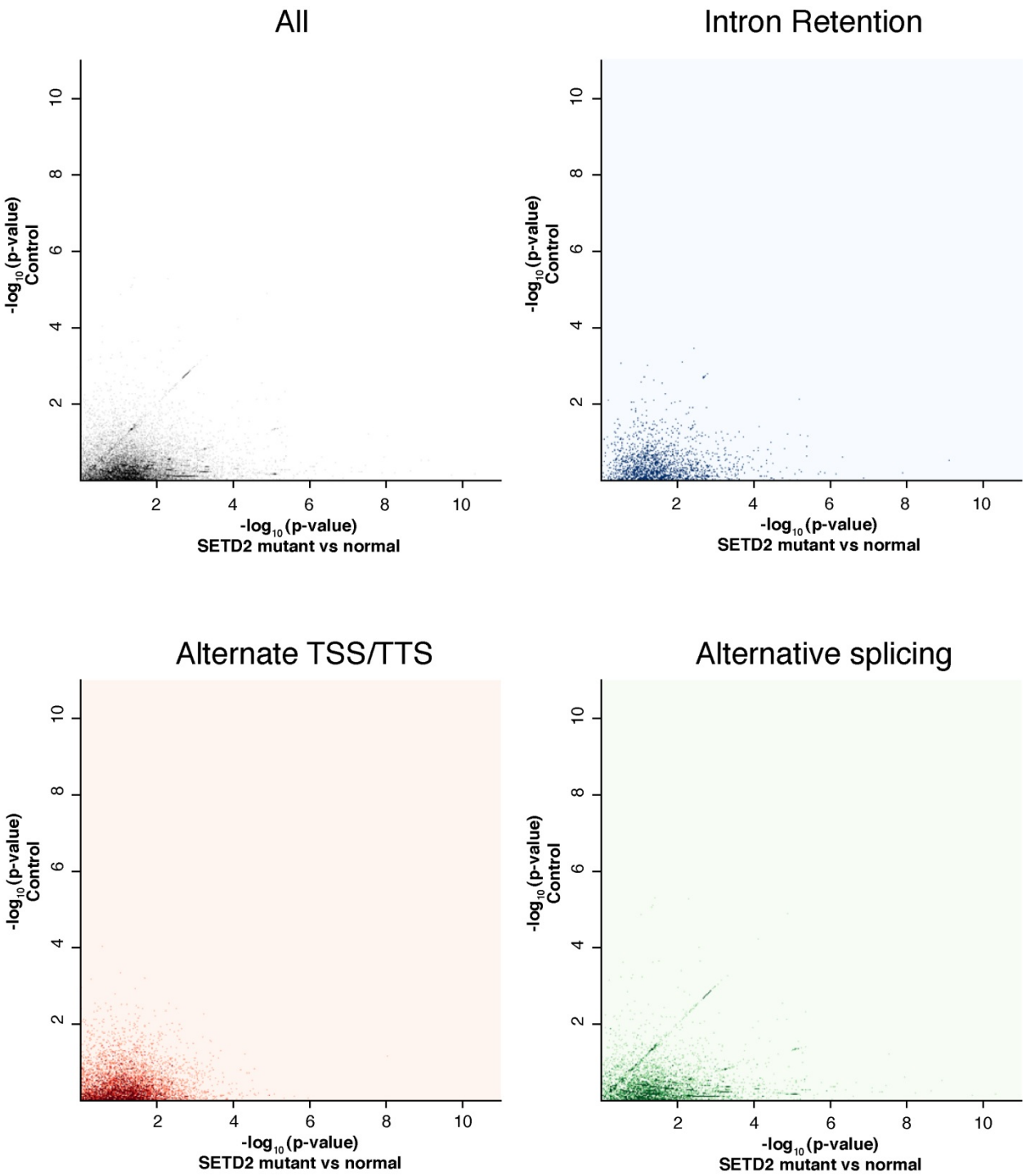
Supplementary Figure 7. GAPDH exhibits low intron retention in H3K36me3-deficient tumors. For *GAPDH*, a gene not exhibiting increased intron retention in H3K36me3 deficient tumors, intron retention scores, genic coverage (calculated with both intron and exon reads), and exon coverage (calculated only with exonic reads) are provided for two H3K36me3 deficient tumors (red) and two H3K36me3 normal tumors (black).



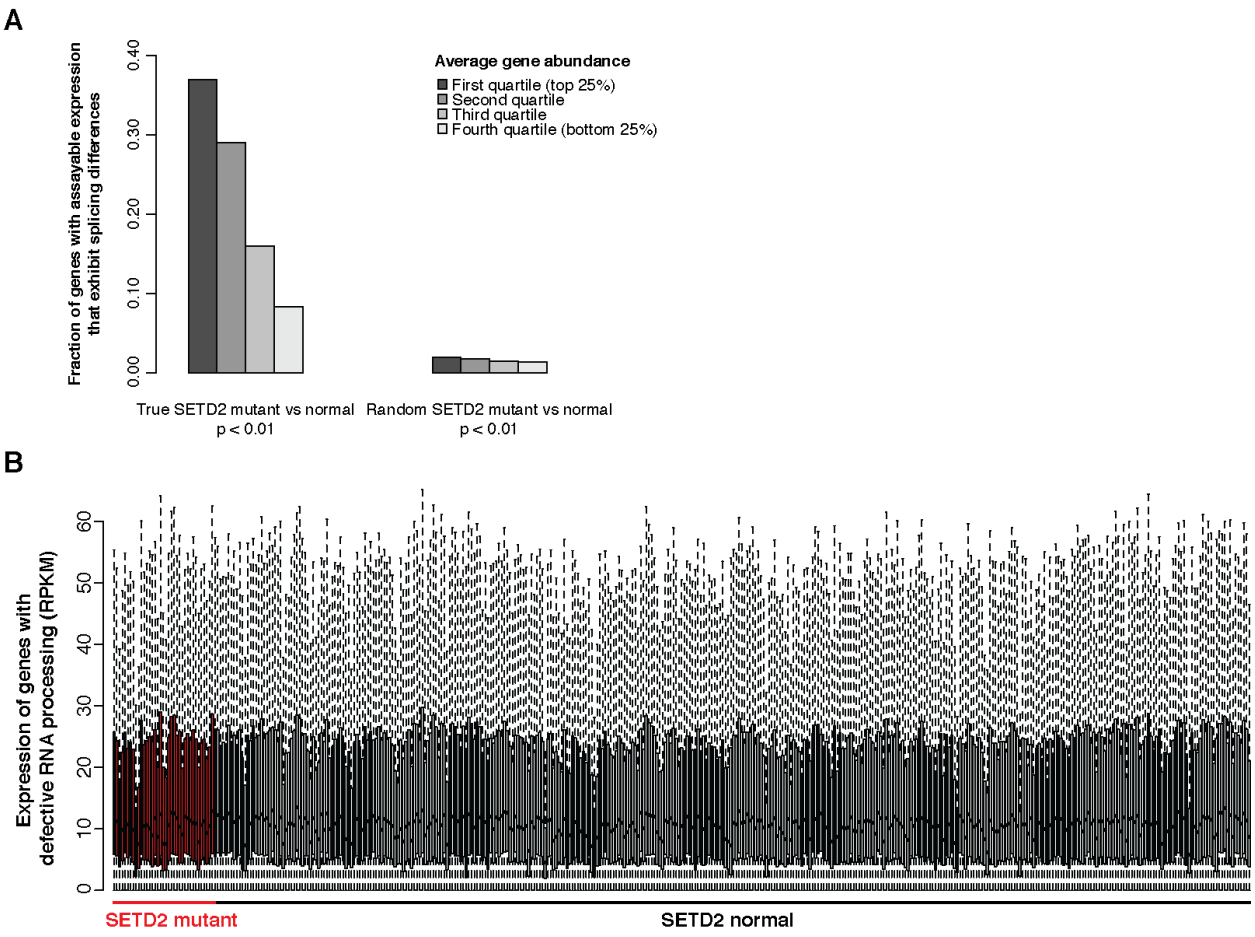
Supplementary Figure 8. Enriched ontologies among genes with increased intron retention. Genes (n=2999) exhibiting increased intron retention between in H3K36me3-deficient tumors were assessed for associated ontologies. The most highly enriched terms among the SP_PIR_KEYWORDS ontology are presented as the $-\log_{10}$ of the Bonferroni-corrected p-value. P-values were filtered to $p < 1 \times 10^{-10}$. "Alternative splicing" refers to genes previously annotated as exhibiting alternative splicing.



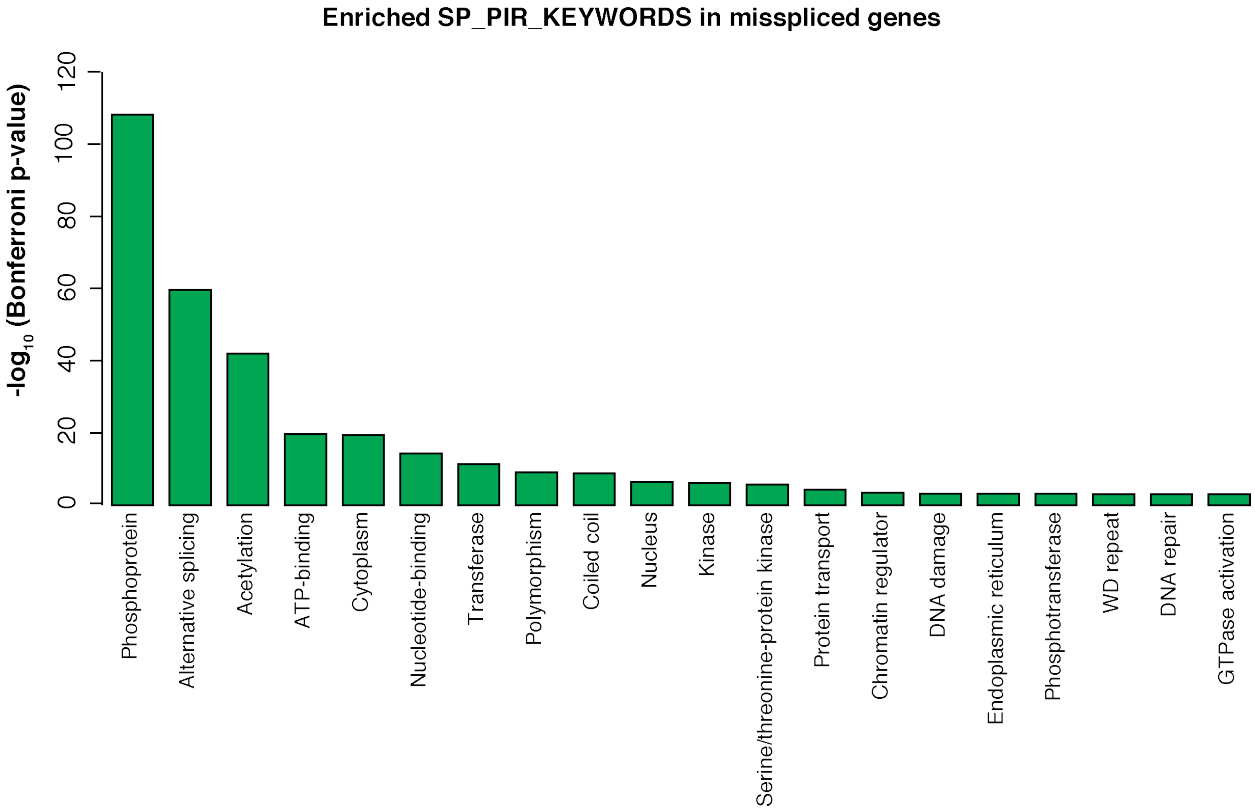
Supplementary Figure 9. **Intron retention in *SETD2*-mutant tumors persists into mature, polyadenylated RNA.** Intron retention scores for 6,546 RefSeq transcripts in polyA$^+$ RNA from the TCGA dataset were averaged across two SETD2 mutant tumors and compared to that of three SETD2 normal tumors. These tumors were selected based on their inclusion and analysis in both datasets. These transcripts were marked by H3K36me3 in normal kidney and contained a site determined by FAIRE to be more nucleosome depleted in H3K36me3-deficient tumors.



Supplementary Figure 10. SETD2-mutant tumors display widespread changes in RNA processing. Significance of the difference in ratios between in *SETD2* mutant and normal tumors (x-axis) are plotted against the scrambled control (y-axis). Combined instances of altered transcript processing (black) can be subdivided as intron retention (blue), alternate transcriptional start or termination sites (red), or alternative splicing (green).



Supplementary Figure 11. Aberrant splicing is preferentially detected in highly transcribed genes. A. RNA abundance for each gene was averaged across all tumors and normal kidney then divided into quartiles. We detected differences in splicing in approximately 38% of the first quartile of genes (top 25% of genes by expression), but only about 8% of genes in the fourth quartile (bottom 25% of genes by expression). **B.** Overall RNA levels (RPKM) for SETD2 mutant (n = 38, marked in red) and SETD2 wild-type tumors (n = 380, marked in black), showing that the expression of genes with defective RNA processing is comparable between these tumor classes.



Supplementary Figure 12. Enriched ontologies among missplicing genes. Genes exhibiting significant splicing differences between SETD2 mutant and normal tumors ($p < 0.003$) in the TCGA cohort were assessed for associated ontologies. The most highly enriched terms in the SP_PIR_KEYWORDS ontology are presented as the $-\log_{10}$ of the Bonferroni-corrected p-value. P-values were filtered to $p < 1 \times 10^{-3}$. "Alternative splicing" refers to genes previously annotated as exhibiting alternative splicing.